

UNCLASSIFIED

Defense Technical Information Center  
Compilation Part Notice

ADP014609

TITLE: Design and Analysis of Entomological Field Experiments

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the Eighth Conference on the Design of  
Experiments in Army Research Development and Testing

To order the complete compilation report, use: ADA419759

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:

ADP014598 thru ADP014630

UNCLASSIFIED

## DESIGN AND ANALYSIS OF ENTOMOLOGICAL FIELD EXPERIMENTS

William A. Brown

Test Design and Analysis Office, Dugway Proving Ground

and

Scott A. Krane

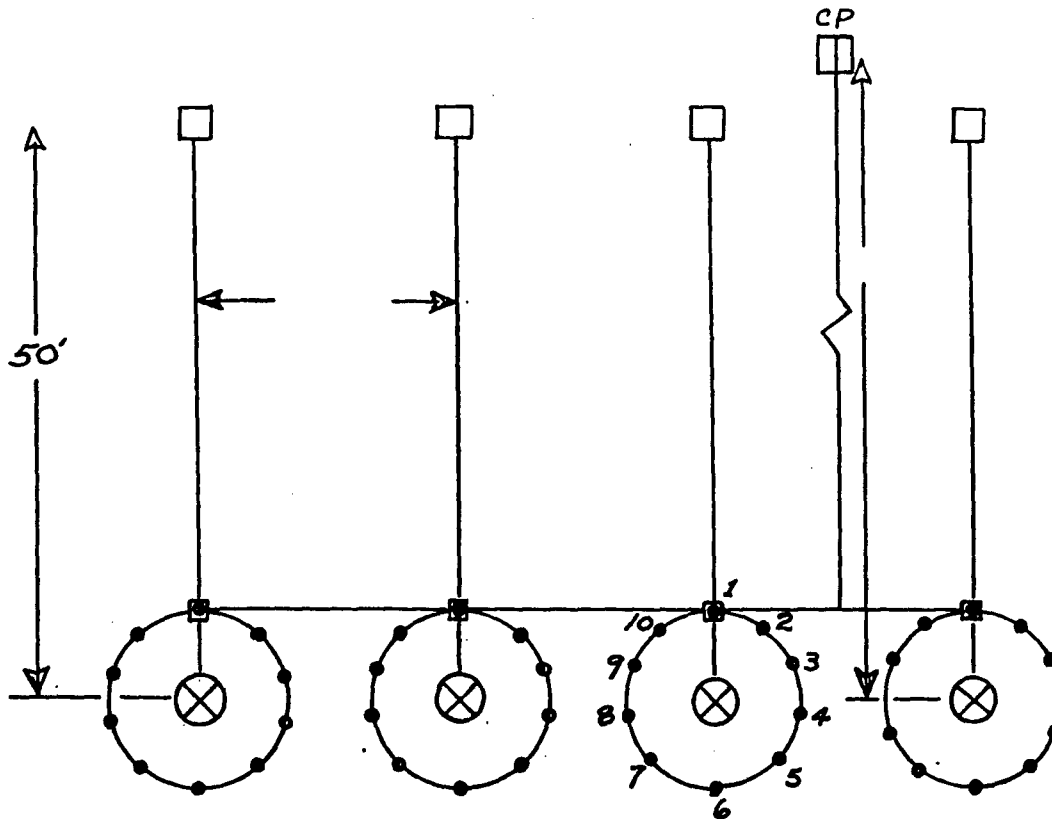
Dugway Field Office, C-E-I-R, INC.

Recently two entomological field experiments were conducted at Dugway Proving Ground. The purpose of the first experiment was to compare the biting propensity of two strains of a species of insect. In each trial, four 15-foot radius circles were scribed, and 10 hosts, randomly selected, were positioned equidistantly along each circumference. The Number 1 position in each circle was oriented to true north. (See Figure 1.) At function time, 100 individuals of the appropriate strain were released at the center of each circle. In two of the circles, the A strain was used; in the other two circles, the B strain. The men were seated on the ground and remained relatively motionless throughout the trial. Sampling consisted of each man recording those bites actually received and entering the total number, for 5-minute intervals, on a data card. Sampling was conducted for 30 minutes following the release unless biting activity continued. In that event, sampling in all circles was extended for additional 5-minute periods until the biting activity had essentially ceased.

Comparisons between strains were thus subject to the variation found among circles. This variation was expected to be appreciably larger than the variation among men on a circle. By the nature of the experimental treatments (strains), however, it was necessary to separate the strains either in space or in time sufficiently that their ranges of biting activity did not overlap. Only in this manner could bites be accurately attributed to one strain or the other. The duplicate circles for each strain represented an effort to partially overcome this inherent insensitivity.

In the analysis of the data, it was considered useful to employ a mathematical model to describe the distribution of the number of bites per host. The simplest model which might conceivably fit the observations is the Poisson, given by:

$$(1) \quad f(x) = e^{-m} m^x / x! \quad x = 0, 1, \dots, n,$$



- Local battery field telephone line; □ telephone
- On-site meteorological sensing station
- Boxed recording instruments
- × Test fixture
- Hosts

Figure 1

where  $x$  is the number of bites received by an individual host,  $f(x)$  is the probability (or relative frequency) of  $x$  bites, and  $m$  is an unknown parameter equal to the "long-run" average number of bites per host. If the individuals of a strain are randomly distributed throughout a given area, and if hosts are equally attractive, then the Poisson model should be appropriate.

Previous studies, however, have indicated that the spatial distribution of insects released in this manner is not random (perhaps being influenced by the wind direction, for example), nor are all hosts equally attractive. As a result of these tendencies, the distribution of bites will be "over-dispersed" relative to the Poisson distribution, i. e., the number of hosts receiving a very large number of bites and the number of hosts receiving a very small number of bites will both be larger than the number predicted by the Poisson model, while the number receiving near-average numbers of bites will be smaller.

One of the simplest and most frequently used "over-dispersed" statistical models is the negative binomial, which has the general term:

$$(2) \quad f(x) = \binom{k+x-1}{x} \frac{p^x}{q^x}, \quad x = 0, 1, 2, \dots, n,$$

where  $q$  equals  $1 + p$ , and  $p$  and  $k$  are unknown parameters. Various rationales may be given for the negative binomial.<sup>1</sup> One of the simplest is that the negative binomial is produced by a mixture of Poisson distributions in which the parameter,  $m$ , varies according to a "gamma" distribution. While no rationale appears to be particularly compelling in the present problem, the relative simplicity of the negative binomial model and the success with which other investigators have applied it to biological data are taken to justify its use, at least as a working hypothesis.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for over-dispersion with respect to a Poisson distribution by the  $\chi^2$  statistic:

<sup>1</sup>Bliss, C. F. Fitting the Negative Binomial Distribution to Biological Data. Biometrics, Vol. 9, (2) pp. 176-196.

$$(3) \quad \chi^2_{n-1} = (n-1)s^2 / \bar{x},$$

where  $n$  is the sample size (number of hosts),  $s^2$  is the sample variance of the number of bites, and  $\bar{x}$  is the sample mean number of bites.<sup>2</sup> The calculated statistic was tested for significance by comparison with the 20 per cent upper tail value of the  $\chi^2$  distribution.<sup>3</sup> If the test did not indicate over-dispersion, the data were subsequently fitted to a Poisson distribution and subjected to a  $\chi^2$  "goodness-of-fit" test. If the test did indicate over-dispersion, the data were fitted by the method of maximum likelihood<sup>4</sup> to a negative binomial distribution, and then subjected to a  $\chi^2$  goodness-of-fit test. All of the above calculations were performed on the IBM 1620 Computer, using a specially prepared FORTRAN program. Fifty-five of 96 sets of 5-minute data showed close agreement with the Poisson distribution. For each of these sets of data, however, the variance was usually larger than the mean, and, consequently, a further comparison with the negative binomial distribution would generally have shown even closer agreement.<sup>5</sup> Therefore, it was decided that, for the purposes of the analysis of variance, an appropriate transformation to stabilize variance for these data would be that derived for the negative binomial:<sup>6,7</sup>

---

<sup>2</sup>Ibid.

<sup>3</sup>For convenience of internal calculation on a digital computer, the 20 per cent upper tail  $\chi^2$  value was obtained from the approximation:

$$\log_e \frac{\chi^2_{n-1}}{n-1} - 1 = -0.038 - 0.452 \log_e (n-1)$$

<sup>4</sup>Fisher, R. A. Notes on the Efficient Fitting of the Negative Binomial. Biometrics, Vol. 9(2), pp. 196-200, 1953.

<sup>5</sup>The Poisson is, in fact, a limiting case of the negative binomial, from which it follows that a negative binomial must fit data at least as well as the Poisson.

<sup>6</sup>Bartlett, M. S. The Use of Transformations. Biometrics, March 1947, Vol. 3(1) pp. 39-52.

<sup>7</sup>Kempthorne, O., Design and Analysis of Experiments. Chapter 8. John Wiley and Sons, Inc., 1952.

$$(4) \quad y = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2}).$$

Figure 2, showing a plot of mean versus variance on log-log paper, illustrates the closer agreement with the negative binomial distribution. The diagonal line represents the square root transformation, appropriate for variance stabilization of Poisson distributed data, and the curved line represents the transformation  $\lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$ , where  $\lambda$  has the value 1.0. (After several guesses of  $\lambda$ , the value of 1.0 was selected since, by eye-fitting, it appeared to reasonably minimize the deviations from the curve. Using the value  $\lambda = 1.0$ , 47 of the data points lie above the curve, and 49 below.) Subsequent analysis of the data of Experiment 1 using a method of Bliss and Owen<sup>8</sup> for the estimation of a common  $k$ , resulted in the estimate

$$k_c = 0.51.$$

Since  $\lambda^2 = (1/k)$ , the value of  $\lambda$  appropriate for this estimate of  $k$  is

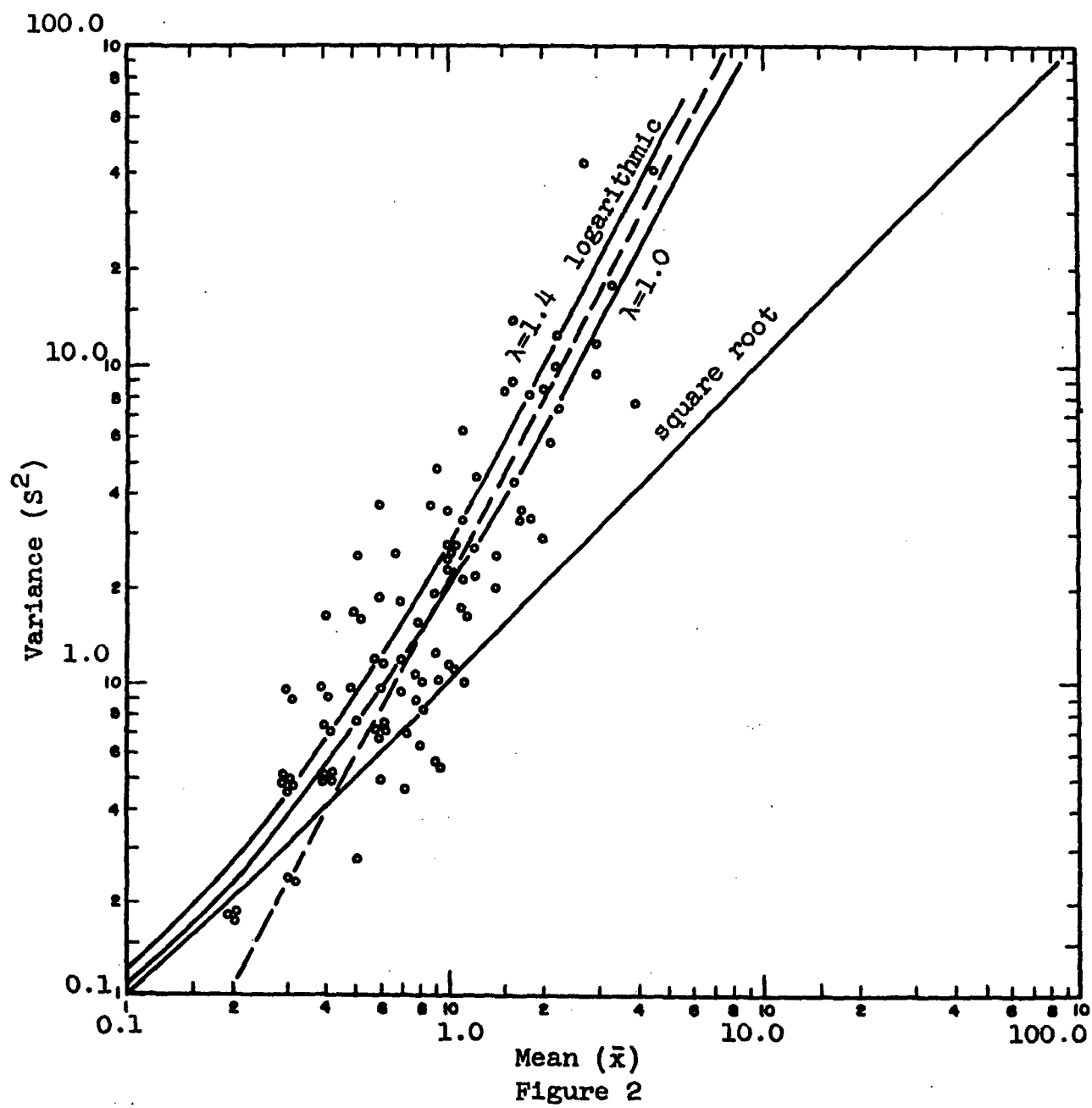
$$\lambda = k_c^{-0.5} = 1.4.$$

As shown in Figure 2, the value of  $\lambda = 1.0$  obtained graphically agrees reasonably well with that estimated by the method of Bliss and Owen. It can easily be seen in Figure 2 that the data follow more closely to the curved line. However, the dashed line indicates that the logarithmic transformation may be as suitable as the inverse hyperbolic sine. Furthermore, analysis of logarithmically transformed data permits interpretations of results, in terms of ratios of treatment effects, while no such interpretation arises directly from the negative binomial transformation. Therefore, separate analyses of variance were performed, using the two transformations.

An analysis of variance, based on the three-way cross classification of trial, strain, and time period, was performed on each of the following four sets of data:

1. The values of  $y = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$ , where  $\lambda$  equals 1.0 and  $x$  is the total number of bites received by a host during a given time period,

<sup>8</sup>Bliss, C. I. and A. R. G. Owen, "Negative Binomial Distributions With a Common  $K$ ", Biometrika 45, pp. 37-58, 1958.



2. The values of  $y = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$ , where  $\lambda$  equals 1.0 and  $x$  is the total number of bites received at a circle during a given time period,

3. The values of  $y = \log (x + 1)$ , where  $x$  is the total number of bites received by a host during a given time period, and

4. The values of  $y = \log (x + 1)$ , where  $x$  is the total number of bites received at a circle during a given time period.

The results of each of these analyses are presented in Table 1.

As shown in Table 1, the results obtained in the four analyses of variance were essentially the same. Each analysis indicated that the total numbers of bites obtained during the six times periods were significantly different, and that no significant difference could be detected between strains. In every analysis, however, Error (a) was relatively large, so that the  $F$  test, comparing strain effects, was undoubtedly insensitive. As mentioned earlier, the insensitivity of the analyses for strain differences follows unavoidably from the design of these trials, in which strain comparisons could only be made between circles (rather than within circles), and, hence, are subject to the greater variability found from circle to circle as measured by Error (a).

The purpose of the second experiment was to compare the dispersal of two strains of a species of insect as measured by their biting activity.

For this experiment, it was greatly desired that the ambient air temperature and windspeed range of an A-B strain pair of trials be as similar as possible. However, because of the small number of men available concurrent testing of the two strains could not be accomplished. Therefore, whenever possible, two trials were conducted each day--one trial using the A strain and the other, following as soon after as practicable, employing the B strain.

In each trial, four concentric circles were used, designated Circles A, B, C, and D with radii equal to 100, 200, 300 and 400 feet. Eight men were positioned equidistantly around each circumference of Circles A, B, and D, and 16 men were positioned equidistantly around the circumference of Circle C. (See Figure 3.) At function time, 1000 individuals of the appropriate strain were released at the center of the concentric configuration, and the men, seated and facing the release point, recorded

Table 1

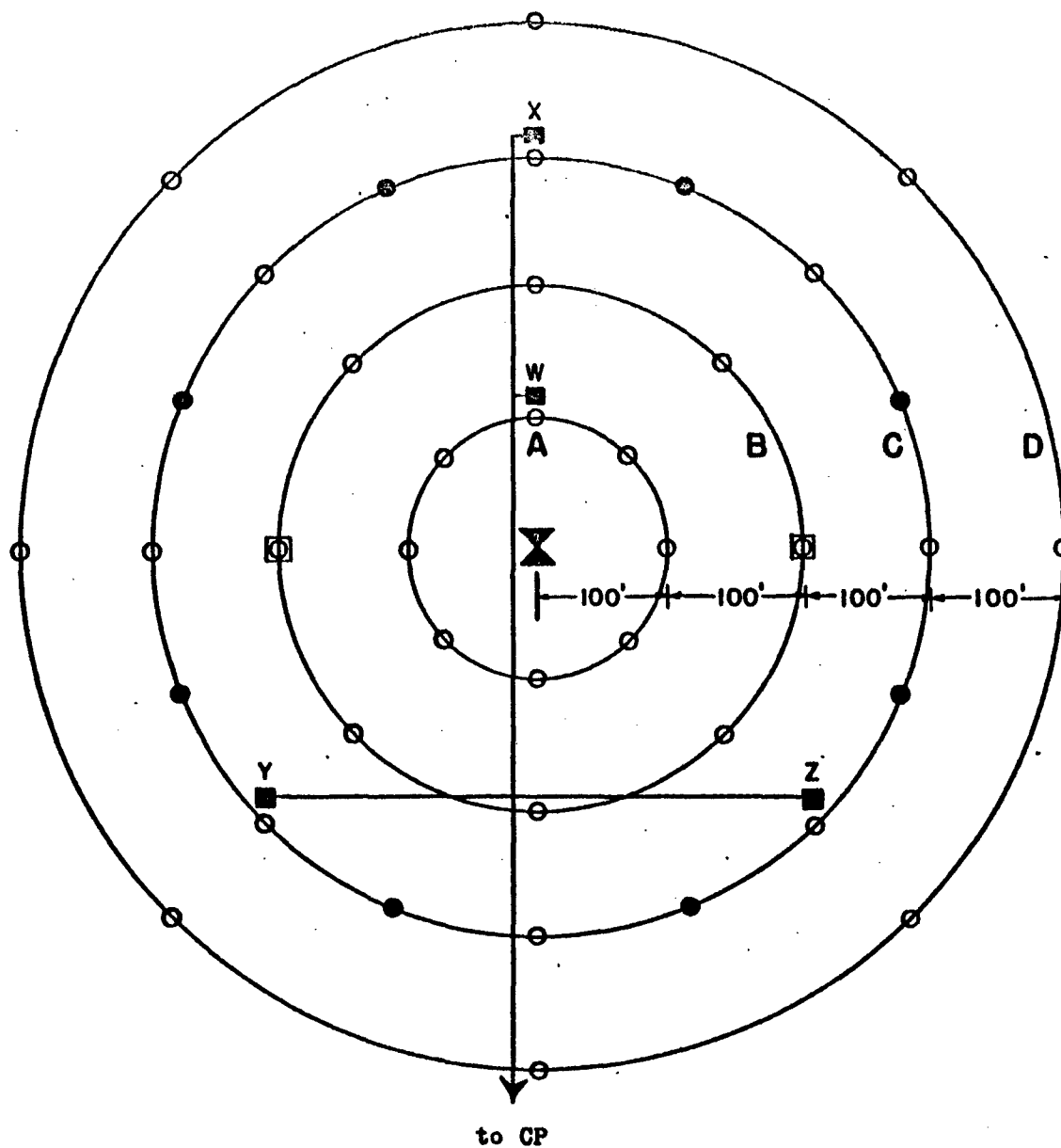
SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMATION*							
		1		2		3		4	
		Mean Square	F Value	Mean Square	F Value	Mean Square	F Value	Mean Square	F Value
Trial, T	5	0.076396		0.054999		0.279571		0.216414	
Strain, S	1	0.240379	3.27	0.240754	2.39	0.894409	3.30	0.944946	2.48
T x S	5	0.017313	0.236	0.25430	0.253	0.063268	0.241	0.095473	0.251
Error (a)	12	0.073442		0.100546		0.270806		0.380320	
Time period, P	5	0.571952	22.8**	0.716051	51.8**	2.073743	60.4**	2.743388	54.1** <sup>3</sup>
P x T	25	0.028863	1.15	0.026287	1.90	0.105622	3.08**	0.101577	2.00* <sup>3</sup>
P x S	5	0.011180	0.445	0.007883	0.570	0.041707	1.21	0.027349	0.540
P x T x S	25	0.016094	0.641	0.019936	1.44	0.058558	1.71	0.075415	1.49
Error (b)	60	0.025108		0.013830		0.034306		0.050667	
Sub-total	143								
Hosts/Circles, C	216					0.147148			
Hosts/P x C	1080					0.029697			
Total	1439								

\*Trials A-3, A-8, and A-9 were omitted from this analysis because of insufficient data.

\*1 denotes  $\lambda^{-1} \sinh^{-1}(\sqrt{x + 1/2})$  transformation of 5-minute total bites received by each host; 2 denotes  $\lambda^{-1} \sinh^{-1}(\sqrt{x} + 1/2)$  transformation of 5-minute total bites received at each circle; 3 denotes  $\log(x + 1)$  transformation of 5-minute total bites received at each circle.

\*\*Significant at the 1.0 per cent level.

\*<sup>3</sup>Significant at the 5.0 per cent level.



- Field telephone
- Field telephone line
- Sampling station
- Augmenting sampler position
- 2-meter wind speed and direction station
- ✕ Release point

Figure 3

biting activity for at least 30 minutes. The sampling procedures were the same as those used for the first experiment.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for "over-dispersion" with respect to a Poisson distribution in the same manner as in the first experiment.

The results indicated that there was nearly always a departure from the Poisson distribution, in the direction of higher variance and "over-dispersion." In addition, 45 of the 70 sets of 5-minute data showed agreement with the negative binomial distribution at a nominal 95 per cent confidence level. Further, from an examination of the plot of the mean versus the variance (see Figure 4), it did not appear that the data would fit any other distribution more consistently. Therefore, it was decided that, for the purposes of the analysis of variance, a suitable transformation to normalize these data would be:

$$y = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2}).$$

After several guesses of  $\lambda$ , and, subsequently, fitting the data by eye to  $\lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$ , it appeared that a reasonable estimate that would minimize the deviations from the curve was  $\lambda = 1.0$ . Using this value, 42 of the data points lie above the curve, and 43 below.

An analysis of variance, based on the four-way cross classification of day, strain, circle, and time period, was performed on each of the following sets of data:

1. The values of  $y = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$  where  $\lambda = 1.0$ , and  $x$  is the total number of bites received during a given time period at Circles A, B, and D, and one-half the total received at Circle C.

The results of these analyses are presented in Table 2.

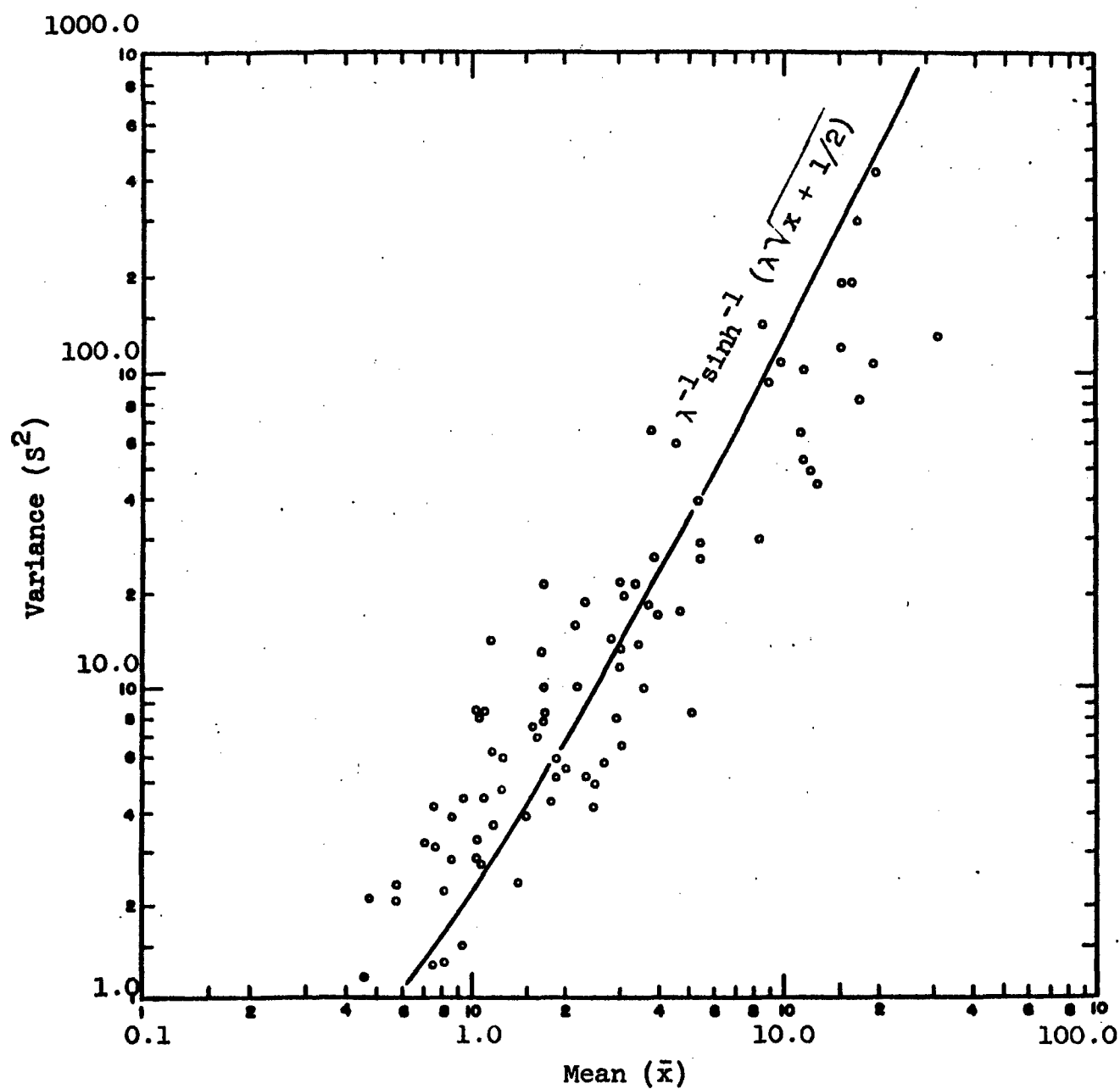
Mean ( $\bar{x}$ )

Figure 4

Table 2: Results of the Analyses of Variance of Transformed Bite Data

SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMED DATA*			
		1		2	
		Mean Square	F Value	Mean Square	F Value
Day, D	1	0.342065		0.331695	
Strain, S	1	0.265038		0.241103	
Error (a)	1	0.271097		0.225376	
Circle, C	3	1.020387	50.5 **	1.199593	71.9 **
C x S	3	0.044502	2.20	0.044751	2.68
Time Period, T	9	0.670518	33.2 **	0.623410	37.4 **
T x S	9	0.115344	5.70**	0.108279	6.49**
T x C	27	0.044332	2.19**	0.043864	2.63**
T x C x S	27	0.011870	0.587	0.014230	0.853
Error (b)	78	0.020225		0.016677	
Total	159				

\*1 denotes the data resulting from the  $\lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$  transformation of total bites received at a circle during a given time period; and 2 denotes the data resulting from the  $\lambda^{-1} \sinh^{-1} (\lambda \sqrt{x + 1/2})$  transformation of total bites received during a given time period for Circles A, B, and D, and one-half the total bites received at Circle C.

\*\* Significant at the 1.0 per cent level.

As shown by the F values in Table 2, the second analysis, adjusting for the augmented sampling on Circle C, was the more sensitive. Both analyses, however, showed circle, time period, T x S, and T x C to be highly significant.

Using the second set of transformed data, a further investigation of circle, T x S, and T x C was made in the following way. From the analysis of the transformed values:

$$(5) \quad y_{ijk} = \lambda^{-1} \sinh^{-1} (\lambda \sqrt{x_{ijk} + 1/2}), \quad \lambda = 1.0$$

$$= \sinh^{-1} \sqrt{x_{ijk} + 1/2},$$

where  $x_{ijk}$  equals the total number of bites received by the  $i$ -th host at the  $j$ -th circle during the  $k$ -th time period. The mean values,  $\bar{y}_{jk}$ , of the transformed variables were obtained. These mean values are related to the estimated true average number of bites received at the  $j$ -th circle during the  $k$ -th time period,  $\hat{m}_{jk}$ , by

$$(6) \quad y_{jk} = \sinh^{-1} \sqrt{\hat{m}_{jk} + 1/2}, \text{ hence}$$

$$\hat{m}_{jk} = (\sinh \bar{y}_{jk})^2 - 1/2.$$

Relationships between  $\bar{y}_{jk}$  and circle radius  $R_j$ , were sought. The best simple relationship found was:

$$(7) \quad \bar{y}_{jk} = a - b \log_e R_j,$$

where  $a$  and  $b$  are regression constants determined by the method of least squares.

$$(8) \quad \text{Then,}$$

$$\hat{m}_{jk} = [\sinh (a - b \log_e R_j)]^2 - 1/2,$$

$$(9) \quad = [1/2(e^{a-b \log_e R_j} - e^{-a+b \log_e R_j})]^2 - 1/2,$$

$$(10) \quad = [1/2(e^a e^{-b \log_e R_j} - e^{-a} e^{b \log_e R_j})]^2 - 1/2$$

$$(11) \quad = \left[ \frac{1}{2} (e^{aR_j - b} - e^{-aR_j b}) \right]^2 - 1/2,$$

which can be approximated by:

$$(12) \quad \hat{m}_{jk} = \frac{2a}{4R_j 2b} - 1/2,$$

since  $e^{-a(R_j b)}$  is small relative to  $e^{a(R_j - b)}$ .

For each 5-minute time period, the transformed data were summed with respect to circle and strain. These values were then fitted to the above regression model, and the average values of the various  $a$ 's and  $b$ 's determined. Subsequently, for each strain, the true average number of bites was estimated for each circle during the various time periods. These latter values are presented in Table 3.

Table 3: Estimated True Average Number of Bites of A and B Strain at the Various Circles During Given Time Periods.

STRAIN	TIME PERIOD (Minutes)	ESTIMATED TRUE AVERAGE NUMBER OF BITES AT INDICATED CIRCLE DURING GIVEN TIME PERIOD			
		Circle A (100 feet)	Circle B (200 feet)	Circle C (300 feet)	Circle D (400 feet)
A	0 - 5	123	15	4	1
	5 - 10	145	26	10	5
	10 - 15	119	28	12	7
	15 - 20	120	29	12	7
	20 - 25	82	20	8	4
	25 - 30	39	15	8	5
	30 - 35	22	10	6	4
	35 - 40	16	8	5	4
	40 - 45	8	6	5	4
	45 - 50	6	4	4	3
B	0 - 5	99	25	11	6
	5 - 10	132	35	16	9
	10 - 15	99	32	16	10
	15 - 20	59	21	12	8
	20 - 25	39	19	13	10
	25 - 30	19	9	5	4
	30 - 35	9	6	5	4
	35 - 40	2	2	2	2
	40 - 45	4	2	1	1
	45 - 50	0	0	0	0

As shown in Table 3, the expected number of A strain bites at Circle A during each of the various time periods is greater than that for B strain; however, the difference, in general, is not appreciable. At Circles B, C, and D, there appears to be no important difference between the number of bites. It was, therefore, concluded that the spatial dispersion of the two strains was comparable, as indicated by the nonsignificance of the C x S interaction.

The significance of the  $T \times S$  interaction indicates a difference between strains with respect to the temporal dispersion characteristics. Generally speaking, the biting activity of strain B appeared to exhibit a more pronounced peak in time and a slightly earlier decline.

Unfortunately, no satisfactory model for the characterization of the biting activity as a function of time has been found by the authors. It is hoped that such a model may yet be developed.